



# 2017 생물통계학 특론

Youngjun Na, PhD



# Youngjun Na

Email: [ruminoreticulum@gmail.com](mailto:ruminoreticulum@gmail.com)

Github: <https://github.com/YoungjunNa>

Major: animal science, ruminant nutrition

Hobby: statistics



## 우리는 이미 통계를 곳곳에서 이용하고 있다. 알게 모르게.

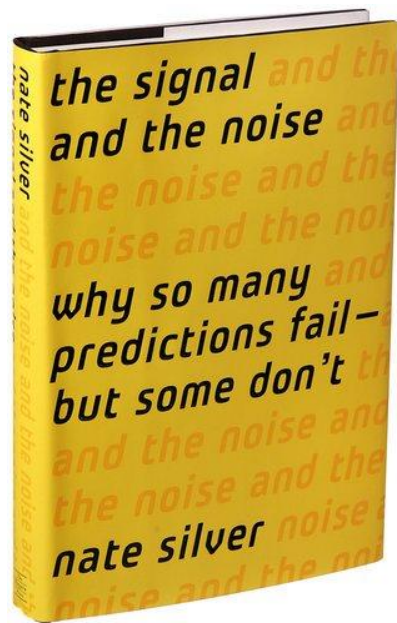
- “우리 목장의 평균 유량은 35kg 입니다” → **현재 상태**를 이야기
- “저번달에 비해 산란률이 10%가 감소했습니다” → **과거와 비교**
- “이 그래프 패턴을 보니 이번 명절 돼지값이 1,000원 정도 증가할 것으로 예상됩니다” → **미래를 예측**

# 불확실한 세계에서 통계의 역할:

## 1. 소음 사이의 신호를 찾아라

“신호는 진리다. 소음은 우리가 진리에 다가서지 못하게끔  
우리의 정신을 산만하게 한다.”

\_Nate Silver



# 불확실한 세계에서 통계의 역할:

## 2. 안경

명사 [a:n.gjən]

어원: 한자 眼鏡

1. 눈의 기능이 나쁜 사람이나 또는 심한 빛이나 먼지로부터 눈을 보호하기 위해 두 개의 유리알을 갈아 귀에 걸쳐 쓸 수 있게 만든 것.





# 데이터 분석의 유형(EDA vs. CDA)

## EDA, exploratory data analysis (탐색적 데이터 분석)

- 데이터의 특징과 구조에 대한 탐구
- 인사이트의 생성
- 先데이터, 後분석

## CDA, confirmatory data analysis (확증적 데이터 분석)

- 가설의 평가
- 모형 적합도, 가설검정, 신뢰구간
- 계획 → 데이터 확보 → 분석

## Reference

1. Tukey, J. (1962). "The future of data analysis". Annals of Mathematical Statistics, 33. 1–67.
2. 허명희. (2015). "데이터 분석의 철학과 과학성". R user conference 2015.



## EDA (탐색적 데이터 분석)

- 기존의 데이터를 들여다 보는 것
- 빅데이터는 EDA의 영역 (기계학습 등)
- EDA를 통해 얻은 insight → 가설 → CDA를 통한 검증



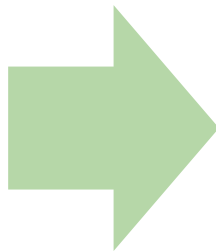
## CDA (확증적 데이터 분석)

- 설계된 실험(designed experiment)을 통해 가설을 검증하는 것 → 생물통계학
- EDA에서 얻은 insight는 수많은 소음에 묻혀 있다
- 통제된 실험 조건(소음의 차단)에서 알고자 하는 처리의 효과(신호)를 알 수 있다



통계, 어려워하지 마세요.

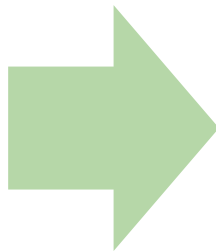
$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$$



**=average(A1:A4)**

통계, 어려워하지 마세요.

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$



**=stdev(A1:A4)**

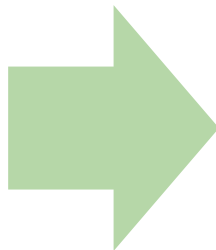
통계, 어려워하지 마세요.

a.  $weight_i = \frac{1}{variance_i} = \frac{1}{SE_i^2}$

b.  $\overline{ES} = \frac{\sum(weight_i ES_i)}{\sum weight_i}$

c.  $SE_{\overline{ES}_w} = \sqrt{\frac{1}{\sum weight_i}}$

d.  $95\%CI = ES \pm 1.96se$



**metacont(n1,m1,s1,n2,m2,s2,  
sm="SMD", study, response)**



# 통계, 어려워하지 마세요.

- 연구자는 연구에 집중하는게 우선
- 통계는 단지 연구를 위한 tool 일 뿐입니다
  - 운전을 하기 위해 자동차의 모든 동작원리를 알 필요는 없음
  - 연구를 하기 위해 통계의 모든 원리를 알 필요는 없음
- 하지만 좋은 연구를 위해서는 적절한 실험 디자인과 통계분석의 기본적인 이해가 있어야 함
  - 우리가 배울 것

## 통계분석 도구



# 우리가 배울 2가지

## 1. EDA

- 통계분석 툴(R studio)을 이용한 탐색적 데이터 분석
- 반응형 보고서 작성

## 2. CDA

- 가설 설정
- 실험 설계
- 통계분석 툴(SAS)을 이용한 통계분석





## 통계분석 단계

	EDA	CDA
1. 데이터의 <b>확보 및 가공</b>	존재하는 데이터 수집	실험을 통한 데이터 수집
2. 데이터의 <b>분석</b>	통계분석	통계분석
3. 데이터의 <b>해석</b>	어떤 insight를 얻을 수 있나?	가설을 증명할 수 있는가?
4. 데이터의 <b>보고</b>	<b>보고서</b> 작성	<b>논문</b> 작성

데이터와 함께 놀아봅시다.



우리가  
갖고 있는  
데이터

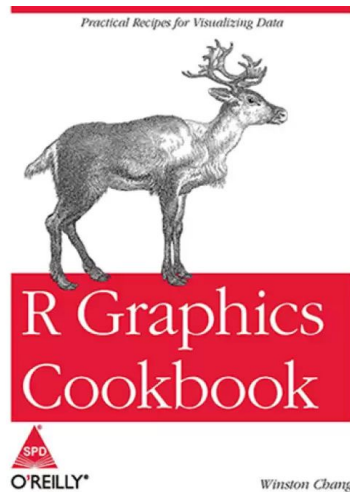


결과물: 논문/보고서



# 교재

1. R graphics cookbook  
([http://book.naver.com/bookdb/book\\_detail.nhn?bid=7307707](http://book.naver.com/bookdb/book_detail.nhn?bid=7307707))
2. Hand out
3. R cheatsheets (<https://www.rstudio.com/resources/cheatsheets/>)





# 숙제

1. R 설치
2. R studio 설치

# 참고: <https://www.rstudio.com/>

**The end**